

## The Art and Science of Protein Purification

### Obtaining protein

In any experiment concerning a protein, it is obviously necessary to obtain material to work with. Most proteins are used by several species. Some proteins are produced only in specific places within an organism, or only at certain times during the lifetime of the organism. The first step is choosing a **source for the protein**, and the species from which to isolate it.

One option, which was historically the only option, and is still a common method, is to find a **natural source**. This means choosing both the organism of choice and organ with the highest-level expression of the protein of choice. Ideally, you want to find a source that is readily available and reasonably inexpensive. In many cases, issues of cost or availability of source material may govern the source species or even the protein of interest. As an example, hemoglobin was heavily studied during the early years of biochemistry because it is both present at very high concentrations in blood, and was easily isolated from blood samples.

Natural sources tend to have problems. The most obvious problems are that many natural sources are difficult to obtain, or may be from less than ideal species, or may produce only very small amounts of the protein of interest. Obtaining a protein from a natural source is especially difficult for human proteins. This is one reason that studies in experimental animals are so popular: it is possible to obtain rabbit livers or cow livers in reasonably large amounts, while obtaining normal human liver is far more difficult.

Natural sources do have the advantage that the protein is present in its normal environment, and therefore (at least at the beginning of the purification procedure) the protein within the source material should exist in its native structure.

A second option, used with increasing frequency, is “**heterologous expression**”, in which the protein of interest is expressed from coding DNA inserted into a host organism. As with natural sources, heterologous expression has advantages and disadvantages. Advantages include the potential for production of the protein in high yield and the possibility of experimentally manipulating the protein sequence prior to synthesis of the protein. Heterologous expression involves a greater investment of time and resources in generating the source in the first place (the host organism must be engineered to produce the protein of interest, and then must be grown in the laboratory). Heterologous expression also involves the possibility that the internal environment of the host organism may lack features required for proper synthesis of the protein. This is especially true for bacterial systems that generally are incapable of incorporating the posttranslational modifications found in many eukaryotic proteins.

Two major types of systems are used for heterologous protein expression: microorganisms and multicellular organisms.

**Bacterial systems** are the most commonly used source of heterologous protein production. Bacteria have several advantages, including high levels of expression,

rapid, inexpensive growth, and amenability to manipulation. Bacterial systems, however, also have some disadvantages. Bacteria are prokaryotes, and their intracellular environment is somewhat different from that in eukaryotes. Differences in intracellular environment and differences in molecular chaperone content may prevent the heterologous protein from folding. In addition, posttranslational modifications (especially glycosylation, but also proteolytic cleavage and other modifications) will either not be performed or may be performed differently in bacteria than in eukaryotic organisms.

**Yeast** grow faster than other eukaryotes, and perform some types of eukaryotic post-translational modifications. However, yeast are more difficult to grow than bacteria, generally result in lower protein expression levels than bacteria, and are difficult to disrupt when attempting to release the expressed protein.

**Higher eukaryotic cells** (such as insect cells or mammalian cells) in culture can also be used to express proteins. The intracellular environments in these cells are much more similar to those found in the natural source. However, cells of higher eukaryotes grow slowly and are expensive to culture. In addition, eukaryotic cells tend to have very low expression levels. They are therefore used largely when the other systems are not likely to work.

Finally, **multicellular organisms** (both plants and animals) can be engineered to overexpress proteins. Engineering multicellular organisms is an expensive and time-consuming process rarely used except for proteins of considerable commercial value.

### **Protein purification**

When attempting to understand how a protein works, it is usually necessary to isolate the protein from other proteins that are present in the tissue. This allows the protein to be studied with some assurance that the results reflect the protein of interest and are not due to other molecules that were originally present in the tissue. In addition, some of the techniques used to study proteins will not yield interpretable results unless the protein preparation is homogeneous. Protein purification is therefore a commonly used biochemical technique.

Most proteins are fairly large molecules. They are smaller than DNA molecules, but they are tremendously large when compared to the molecules typically used in organic chemistry. The three-dimensional structure of most proteins is a consequence of many relatively weak non-covalent interactions. Disrupting this three-dimensional structure, on which the function of the protein depends, is therefore a relatively easy process. Conversely, preventing the loss of the non-covalent structure (and sometimes the covalent structure) is frequently difficult.

Disrupting cellular structure is required to release the proteins from the cell. However, the process has two side effects that may damage proteins: 1) cell disruption typically involves shearing forces and heat, both of which can damage proteins, and 2) cells normally contain proteases (enzymes that hydrolyze other

proteins). In most cells, proteases are carefully controlled; however, disruption of the cell usually also releases the proteases from their control systems, and may allow the cleavage of the protein of interest.

Purification of proteins involves taking advantage of sometimes-subtle differences between the protein of interest and the remaining proteins present in the mixture. Because proteins are all polymers of the same twenty amino acids, the differences in properties tend to be fairly small.

In most cases, current understanding of protein structural and chemical properties is insufficient to allow a purification method to be generated theoretically. The “Art” in the title of this section reflects the fact that development of most protein purification procedures is a matter of trial and error. The table below lists some of the general properties of proteins that can be useful for protein purification, and some of the methods that take advantage of these properties. Each of these general methods will be discussed in some detail below. Note that for any given protein, only some of these methods will be useful, and therefore protein purification schemes vary widely.

<b>Property</b>	<b>Technique</b>
Solubility	Ammonium sulfate precipitation
Charge	Ion-exchange chromatography
Hydrophobicity	Hydrophobic interaction chromatography
Size	Gel-filtration chromatography
Function	Affinity chromatography
Stability	Heat-treatment, pH treatment

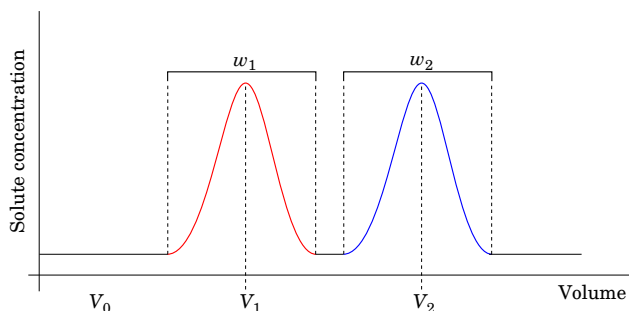
### **Chromatographic methods**

Most protein purification methods involve liquid chromatography. Liquid chromatographic methods employ a column of an insoluble material (the “**stationary phase**”) that can bind molecules based on specific properties common to proteins. A solution (the “**mobile phase**”) containing the mixture of proteins is allowed to pass through the column; the protein of interest may bind to the stationary phase (depending on the properties of the stationary phase and of the protein), while at least some impurities remain in solution and leave the column. The procedure is completed by eluting (*i.e.* removing) the proteins that have bound to the column. Separations are achieved as a result of the **differential partitioning** of the proteins and other solutes between the stationary phase and the mobile phase.

The effectiveness of a chromatographic technique is contingent on a variety of factors, including the properties of the proteins in solution, the properties of the resin used as the stationary phase, the properties of the functional groups covalently attached to the column resin, and the properties of the mobile phase. Adjusting these properties can result in improved purification; making adjustments efficiently requires at least a partial understanding of the theory underlying chromatography.

## Chromatography theory

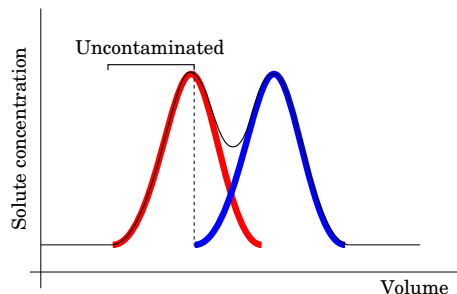
When a solute dissolved in the mobile phase is applied to a column, it will migrate through the column, interacting to a variable degree with the column resin. Eventually, the solute will leave the column. For each solute, it is possible to measure the volume of mobile phase required before the solute leaves the column. A chromatogram for two solutes is shown below. In this chromatography experiment,  $V_0$  is the void volume, the volume of mobile phase external to the resin beads that is experienced by a solute too large to enter the beads.



Assessing the ability of a column to separate different solutes is necessary to decide whether a given column is useful for separating those solutes. Resolution is a measure of the separation observed for two different solutes passing down the column. Resolution is given by:

$$R_s = \frac{2(V_2 - V_1)}{w_1 + w_2}$$

where  $V_1$  is the elution volume of the first peak,  $V_2$  is the elution volume for the second peak, and  $w_1$  and  $w_2$  are the baseline peak widths for the two peaks. The equation implies that, for broad peaks, a larger difference in elution volume will be necessary for high resolution. High resolution is considered to be achieved for  $R_s$  values of 1.0 or greater.  $R_s$  values of  $>1.5$  correspond to baseline separation, in which  $\sim 100\%$  of one solute may be collected without collecting the other solute. The peaks in the graph shown below are incompletely resolved, with portions of the first peak solute contaminated with the solute from the second peak. Collecting the uncontaminated material requires discarding the material from solution containing both solutes. Thus, obtaining high yields of purified material requires high resolution.



Resolution is also given by another equation that looks at the separation slightly differently:

$$R_s = \frac{1}{4} \frac{(1-\alpha)}{\alpha} \sqrt{N} \frac{k_B}{(1+k_B)}$$

In this equation,  $k$  is the retention factor. For each peak, the retention factor is given by:

$$k = \frac{V_1 - V_t}{V_t}$$

This equation introduces a new term,  $V_t$ , the volume of a cylinder the size of the column bed. In calculating  $R_s$ , the  $k$  used is the average of the  $k$  values for the peaks being assessed. For columns in which the solutes bind to the column with high affinity, the elution volume  $V_1$  can be greater than  $V_t$ .

Another important term in the  $R_s$  equation is the number of theoretical plates,  $N$ . A chromatography column can be considered to comprise a stack of “plates”, where each plate is a thin disk of small volume. If a solute is applied to the top of the column and allowed to migrate through the column, the solute will experience a volume of buffer required to elute the solute. For a given column,  $N$  is measured for a solute that does not interact with the resin; it is therefore a property associated with the minimal elution volume for a given column.

$$N = 5.54 \left( \frac{V_1}{w_{1/2}} \right)^2$$

$N$  is dependent both on factors associated with the resin (especially the length of the column and the uniformity and size of the beads, with smaller beads corresponding to higher  $N$ ) and on factors associated with the chromatographic run itself (such as flow rates and the amount of sample loaded).

Columns with high  $N$  are efficient and have tall sharp peaks. Low values of  $N$  correspond low efficiency and short broad peaks.

The final term is the selectivity,  $\alpha$ , which is given by:

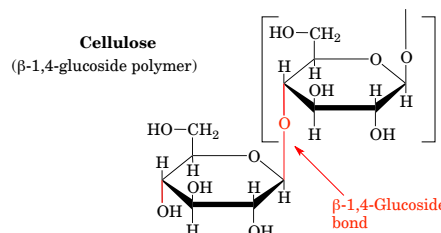
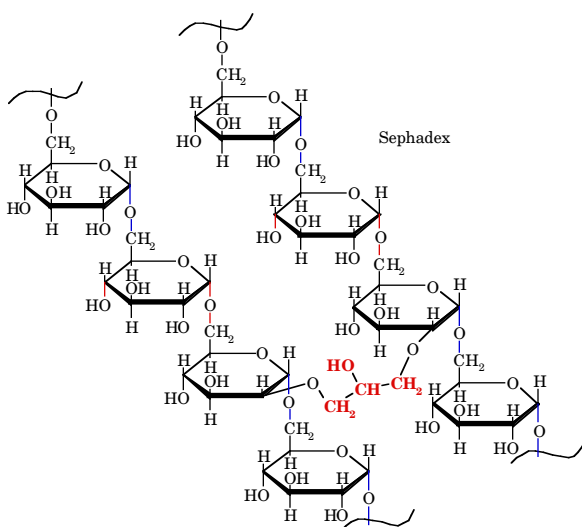
$$\alpha = \frac{k_2}{k_1} = \frac{V_2 - V_0}{V_1 - V_0} = \frac{V_2}{V_1}$$

Selectivity is a measure of the relative binding affinity of the two solutes to the column. An examination of the equation for resolution reveals that higher selectivity is more important than greater  $N$ . Although greater numbers of theoretical plates can be obtained by using a longer column (with resolution increases as the square root of the column length), increasing selectivity usually requires finding a more suitable column resin.

## Resin properties

Suitable resins for protein chromatography should have good flow properties, should be reasonably hydrophilic, and should be porous to provide a large surface area. Most resins used for protein purification are based on carbohydrate polymers, which generally exhibit these properties.

Resin	Carbohydrate	Cross-linking agent
Biogel	Polyacrylamide	Bisacrylamide
Sephacel	Cellulose	Epichlorhydrin
Sephacryl	Dextran	Bisacrylamide
Sephadex	Dextran	Epichlorhydrin
Sepharose	Agarose	none
Superdex	Dextran, Agarose	Highly cross-linked
Superose	Agarose	Highly cross-linked

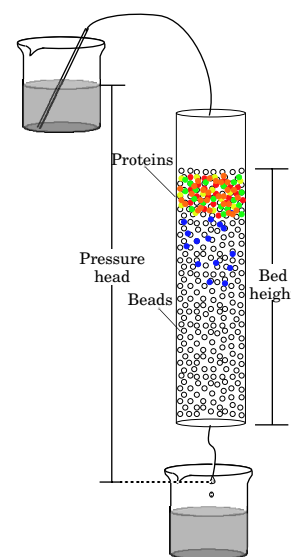


## Pressure

Running a chromatography column requires a method for forcing the liquid mobile phase to flow through the column. The liquid can be forced through the column using gravity or a pump.

Pressure is usually measured either in pounds per square inch (psi) or in pascals (Pa), where 100 kPa = 14.7 psi = 1 atmosphere. In a sealed system using gravity, the pressure head is the vertical distance between the surface of the source buffer and the end of the outflow tubing. Vertical distance corresponds to a potential energy gradient; because most protein liquid chromatography systems use aqueous buffers, a 10 meter vertical distance corresponds to about 100 kPa, although denser buffers will result in higher pressures.

It is apparent that in most laboratories pressures above about 20 kPa are impractical using gravity. As a result a variety of



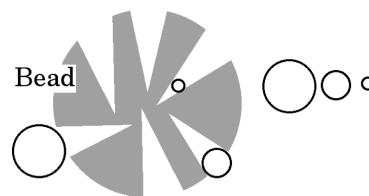
pump systems are used to provide the necessary force. In addition to being capable of producing pressures of up to 50 MPa, pumps can be regulated to allow tighter control of flow rates.

Chromatography columns will be damaged by pressures that exceed the structural strength of the beads. Different resins have different pressure tolerances, with sephadex resins being only able to withstand low pressures. Columns produced using superose or superdex resins generally require pumps, because their flow resistance is too high to allow the use of gravity.

### Gel filtration chromatography

In gel filtration chromatography (also known as size exclusion, gel permeation, or molecular sieve chromatography), molecules are separated based on size. Gel filtration columns are made of porous beads packed into a column. The beads can be polymers of dextran (Sephadex), agarose (Sephacryl and Superose), agarose cross-linked to dextran (Superdex), polyacrylamide (Sephacryl), or related compounds. Different types of beads have somewhat different physical properties that may make them more appropriate for different proteins. Beads with a variety of pore sizes are available; the different pore sizes allow separation of different molecule size ranges. A gel filtration chromatography column can be calibrated by using molecules of known size; the gel filtration column can then be used to determine the size of an unknown protein. In contrast to SDS PAGE (discussed below), which measures the molecular weight of denatured protein, gel filtration measures the size of the native protein.

As a solution containing molecules of varying sizes passes through the column, the molecules distribute between the inside and outside of the pores depending on their size. Molecules too big for the pores are totally excluded, and elute from the column first. Smaller molecules fit in the pores, and therefore elute later. The elution volume for a molecule is thus inversely related to the size of the molecule. This process can be treated mathematically by considering the behavior of solutes within the column.



The “void volume”  $V_0$  is the elution volume of a molecule so large that it is totally unable to enter any of the pores. The void volume therefore represents the minimum possible elution volume for any molecule. The column volume  $V_t$  is the volume of cylinder that contains the resin bed.  $V_t$  is the sum of the void volume  $V_0$ , the volume of solid beads  $V_g$ , and the internal volume of the beads  $V_i$ . The amount of solute inside the beads,  $m_i$ , is given by:

$$m_i = \sigma V_i c$$

where  $\sigma$  is the partitioning coefficient, and  $c$  is the local concentration of solute outside the bead. The partition coefficient varies, with  $\sigma = 0$  for molecules unable to enter the bead at all,  $\sigma < 1$  for molecules too large to enter some pores,  $\sigma = 1$  for

molecules readily able to enter the pores, and  $\sigma > 1$  for molecules that interact with the beads. (In gel filtration, the bead is assumed to be inert; if  $\sigma > 1$ , this assumption is not valid, and therefore the solute will migrate more slowly than it should based on its size.)

Assuming that the  $\sigma$  is not greater than 1,  $\sigma = V_p/V_i$ , where  $V_p$  is the volume accessible to the solute. If  $\sigma = 0$ , the solute runs at the void volume. If  $1 \geq \sigma > 0$ , then the solute will see the additional volume  $V_p = \sigma V_i$ .

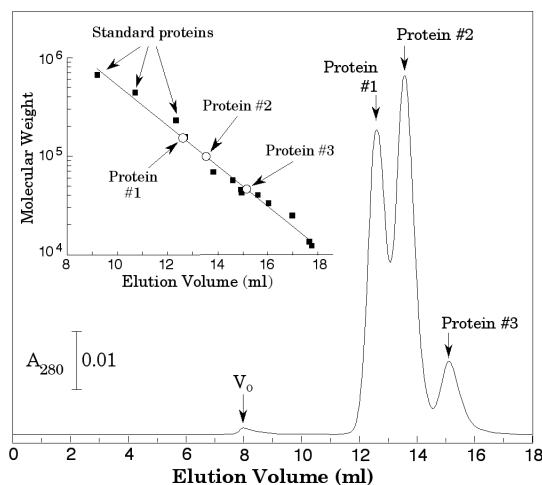
The volume will then be  $V_e = V_0 + \sigma V_i$ . This equation allows  $\sigma$  to be calculated, since  $V_i$  can be measured. ( $V_i$  is the volume of buffer required to hydrate the column; for many beads,  $V_i \approx V_t$ .) Experimental analyses have indicated that  $\sigma$  is a function of the Stokes radius (a property of diffusion of a molecule proportional to the size of the molecule):

$$\sigma = A \log r_h + B$$

where  $r_h$  is the Stokes radius, and A and B are constant properties of the column. If the column is calibrated with molecules of known  $r_h$ , the  $r_h$  of the new molecule can be estimated.

In most cases, the log of the molecular weight is used instead of the Stokes radius. While molecular weight is a more intuitively useful value, differences between the shape of the solute being studied and the standards used to calibrate the column can result in inaccurate molecular weight determination.

The figure (below) shows a gel filtration chromatogram separating three proteins. The inset shows a standard curve (a plot of log of molecular weight *versus* elution volume) generated using a number of proteins (black squares); it also shows the migration positions of the three unknown proteins plotted on the standard curve line. From this graph, it is possible to estimate the molecular weight of the unknown proteins. In this case, the three proteins shown were known to be the only proteins present in the solution.





### **Gel filtration experimental comments**

Gel filtration chromatography can be performed using gravity (especially for large columns used for protein purification), or using low pressure or high pressure pumps (especially for analytical purposes). HPLC columns have high resolution, and can usually be run fairly rapidly.

Gel filtration beads are characterized by their composition. They are also characterized by their exclusion limit, which is the size of the smallest molecule incapable of entering the beads. Beads with a variety of compositions and pore sizes are available. For example, a Sephadex G-100 bead is made of a dextran polymer. The “100” refers to the fact that a column of the beads should be able to separate molecules of up to 100 kDa; molecules bigger than 100 kDa will be completely excluded and will run at the void volume.

Some columns are more efficient than others (recall the discussion of  $N$ , above); efficient columns can tolerate larger sample volumes than inefficient columns of the same general properties.

Certain problems are inherent in the process of gel filtration. The protein sample should be applied to the column in the smallest reasonable volume to ensure good separation. The molecules do not form tight associations with the column, and if loaded in a large volume, will elute in a large volume with little separation between large and small molecules. Even if the sample is applied as a narrow band to the column, the turbulence associated with passing the mobile phase through the column results in a broadening of the bands of the applied molecules as they travel through the column. The broadening of the bands works against the purification of the individual components of the sample, since the broad peaks can partially overlap one another. Unfortunately, some samples may be difficult to concentrate sufficiently to make the method optimal. The maximum volume that can be applied to a column varies somewhat with bead type (*e.g.*, about 2 to 3% of the bed volume for Sephadex, and about 0.5 % of the bed volume for Superdex).

Another problem with gel filtration is that proteins may interact with the column matrix. This interaction will retard the elution of the protein and the apparent molecular weight will be lower than expected. To prevent interaction of the protein with the matrix the gel filtration should be run under moderate ionic strength to prevent ionic interactions between the proteins and the column matrix.

Separation on gel filtration columns increases in proportion to the square root of the column length. Longer columns therefore result in higher resolution. Unfortunately, both the cost of the column and the time the column takes to run increase in direct proportion to the length of the column.

Separation is inversely proportional to flow rate; thus, improved separation typically requires longer chromatography times. The maximum flow rate for a particular separation on a particular column can be estimated from theoretical considerations, but usually has to be determined empirically (*i.e.* try a flow rate, and see if the results are acceptable).

### **Experimental uses of gel filtration chromatography**

Gel filtration can be used as a purification technique. In most cases, gel filtration is used as a later step in the purification procedure, because it is a relatively low resolution technique, and because it cannot tolerate large sample volumes. However, gel filtration can be used as a desalting technique at any time during the purification.

As mentioned above, gel filtration is used as an analytical technique to measure the molecular weight of proteins in solution. In performing these experiments, it is worth examining the migration of different standard molecules; if molecules of known Stokes radius do not fit the calibration curve, it may be necessary to modify the mobile phase either to improve stability of the proteins or to prevent interactions with the resin.

In some cases, molecules will be observed with  $\sigma < 0$ ; this generally implies that the column has channels that allow the molecules to bypass portions of the resin, and therefore the column must be repoured.

In theory, the peaks observed for migrating proteins should be gaussian curves, because the distance from the location of maximum concentration (due to turbulence and diffusion effects) should follow a probabilistic distribution. Inhomogeneities in the column will result in peaks having non-gaussian shapes, and may necessitate repouring the column.

Gel filtration is also used to assess binding interactions between proteins and between proteins and small molecules. For high affinity interactions, the protein and ligand should separate into distinct peaks, which allows the measurement of the bound to free distribution. For lower affinity interactions, dissociation of bound material will occur during chromatography, and the bound ligand will appear as a non-gaussian peak with a long tail. In principle, the shape of the peak can be used to measure the binding affinity. In practice, this requires a high quality column and tends to present several technical problems.

The tubing after the end of the column should have a minimal length to prevent mixing phenomena that will reduce the apparent efficiency. This is true for all types of chromatography.

### **Ion exchange chromatography**

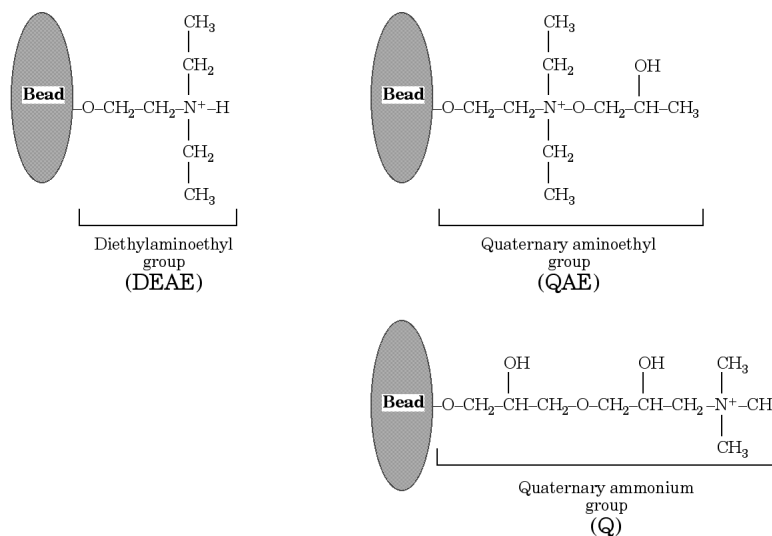
Proteins are charged molecules. They will therefore bind to other molecules of opposite charge. Ion exchange columns are produced by covalently attaching charged molecules such as diethyl-aminoethyl groups to insoluble carbohydrate resins. In many cases, small differences in charge can result in significant separations on ion exchange columns. Ion exchange columns are typically loaded at low ionic strength, and the protein removed by raising the ionic strength.

Ion exchange is a very powerful technique. Most ion exchange resins have a high capacity for protein binding, and can yield high-resolution separations. Ion exchange chromatography is therefore a major tool used for protein purification.

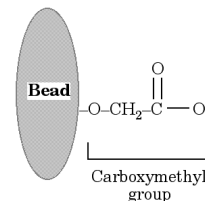
Early ion exchange resins were hydrophobic and non-porous; high surface charge density and hydrophobic resin tended to denature proteins. In addition, had low capacity for protein binding due to inability to bind except at surface. Current resins are porous and hydrophilic carbohydrate polymers. The resins used are comprised of the same cross-linked carbohydrates as those used for gel filtration, except that the resins have functional groups covalently attached to the hydroxyl groups of the carbohydrates. The levels of substitution can vary depending on the type of resin. Most of the resins used have large pores, with exclusion limits above  $10^6$  Da.

**Strong** ion exchangers are ionized over a wider range of pH values than are **weak** ion exchangers (the strength does *not* refer to solute binding affinity). Strong exchangers can be used between pH 2 and 12 (outside this range tend to damage the resin). Weak exchangers are charged only at narrower ranges of pH.

**Anion exchangers** have positively charged functional groups. The structure of the DEAE group, a weak anion exchanger that is ionized between pH 2 and 9, and of two strong anion exchangers based on quaternary amino groups.



Cation exchange resins have negative charges. One commonly used weak cation exchange functional group is the carboxymethyl (CM) group, which is charged between pH 6 and 10. Strong cation exchange resins have sulfonic groups on linkers attached to the resin.



Ion exchange chromatography is pH dependent. Most proteins bind to anion exchange resins at pH above their pI, and to cation exchangers at pH below their pI.

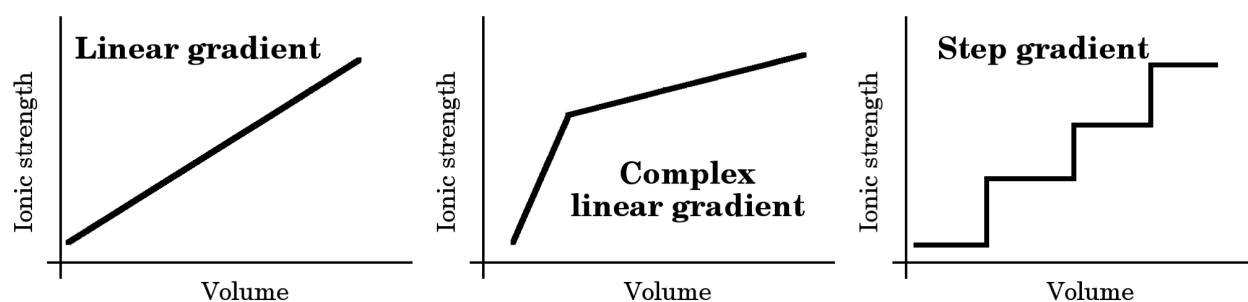
Unlike in gel filtration, sample volume used for ion exchange chromatography is essentially irrelevant, because the sample binds to the resin. The only important parameters are the ionic strength of the sample solution, and the amount of protein relative to the capacity of the column. Ideally, the sample should correspond to a

maximum of 20% of the binding capacity of the column. If the column has excess capacity, it is larger, and therefore more expensive than necessary. If the column capacity is insufficient, the protein may elute from the column in a less controlled fashion, which can reduce the effective resolution of the column.

## Gradients

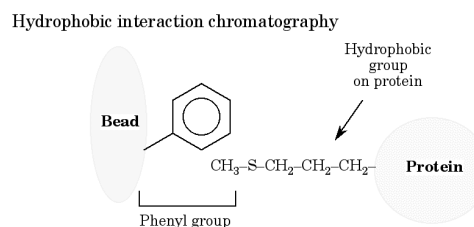
In order to elute protein bound to an ion exchange column, it is usually necessary to increase the ionic strength of the mobile phase, generally by adding increasing amounts of a salt such as NaCl or KCl. Because different proteins will elute at different ionic strength, the rate at which the ionic strength increases has a considerable effect on the resolution.

In designing a purification method, it is advisable to use a simple linear gradient to increase the ionic strength. This allows the resolution of the protein of interest from contaminants to be assessed. Once the ionic strength required to elute the protein of interest is known, it may be worth modifying the gradient to improve either the resolution or the length of time required for the chromatography. Some examples of gradients are shown below.



## Hydrophobic interaction chromatography

Proteins contain hydrophobic amino acid side chains, some of which are exposed at the surface of the protein. Proteins will therefore often bind to other hydrophobic molecules. Hydrophobic interaction columns are produced by covalently attaching hydrophobic molecules such as short acyl chains or phenyl groups to insoluble carbohydrate resins.



Hydrophobic interaction chromatography is similar to **reverse phase chromatography** in concept. However, reverse phase chromatography generally employs different support materials (especially silica), the support is typically more highly substituted, and reverse phase resins usually have larger non-polar groups attached. Perhaps the most common reverse phase functional group is a linear C18 chain; in hydrophobic interaction chromatography, the functional groups are somewhat less hydrophobic (using C4 or C8 chains or the phenyl groups shown above). Because of the more hydrophobic resins, reverse phase chromatography frequently requires organic solvents as mobile phases, while Hydrophobic interaction chromatography uses aqueous buffers.

In general, the hydrophobic effect is strongest under high ionic strength conditions; hydrophobic interaction columns are therefore typically loaded at high ionic strength, and the protein removed by lowering the ionic strength (thus, these columns are the opposite of ion exchange columns). The effect of different ions in altering the hydrophobic effect appears to follow the Hofmeister series, shown below in terms of decreasing salting out effect.

Anions:  $\text{PO}_4^{3-} > \text{SO}_4^{2-} > \text{CH}_3\text{COO}^- > \text{Cl}^- > \text{Br}^- > \text{NO}_3^- > \text{ClO}_4^- > \text{I}^- > \text{SCN}^-$   
Cations:  $\text{NH}_4^+ > \text{Rb}^+ > \text{K}^+ > \text{Na}^+ > \text{Cs}^+ > \text{Li}^+ > \text{Mg}^{2+} > \text{Ca}^{2+} > \text{Ba}^{2+}$

The source of the Hofmeister series is poorly understood. Part of the effect appears to be related to alterations in water structure; salts that result in increasing water surface tension in general exhibit greater salting out effects. However, the ability to increase water surface tension does not follow the Hofmeister series exactly. The salts shown below are ordered in terms of decreasing ability to increase surface tension of water.

$\text{Na}_2\text{SO}_4 > \text{K}_2\text{SO}_4 > (\text{NH}_4)_2\text{SO}_4 > \text{Na}_2\text{HPO}_4 > \text{NaCl} > \text{LiCl} > \text{others} > \text{KSCN}$

pH also tends to affect the binding of proteins to hydrophobic interaction resins, with lower pH usually being associated with stronger binding. The optimum pH for a HIC separation is usually determined empirically.

Running ion exchange followed by hydrophobic interaction chromatography takes advantage of the fact that ion exchange columns are eluted by high ionic strength, and HIC columns are loaded at high ionic strength. While running HIC first followed by ion exchange also takes advantage of the salt concentration alterations, the performing the HIC last yields the protein with much smaller amounts of contaminating salt.

### **Batch adsorption**

With ion exchange and HIC resins, it is possible to add the resin to the protein solution. The advantage is that the resin can bind the protein rapidly, and in large amounts, followed by simple filtration to remove the unbound material. The disadvantage is that the high-resolution separations achievable using chromatography are not possible, and in most cases, the yield of bound protein is lower.

### **Affinity Chromatography**

Some proteins bind to small molecules or to other proteins with high affinity. If so, it may be possible to cross-link the small molecule or other protein to a chromatography resin. Separating your protein then involves loading the protein mixture; proteins that do not bind are eluted by washing with buffer.

Affinity chromatography can be very efficient and proteins can be purified several thousand fold in a single step.

Many expression systems now take advantage of affinity chromatography: you force the bacteria to produce your protein as part of a larger polypeptide, with the additional sequence containing a domain that will specifically bind to an affinity column.

Examples of affinity chromatography:

**ADP-agarose** – many proteins that bind NADPH will also bind adenosine-bisphosphate (usually, incorrectly, abbreviated as ADP). Cross-linking adenosine-bisphosphate to a resin such as agarose results in a column that NADPH-dependent enzymes will bind preferentially.

**His tag** – A sequence of six histidines will bind metals with high affinity. Adding six histidines at the N- or C-terminus of your protein coding region often has little effect on the protein structure, and will usually allow binding to a column that has metal ions (usually nickel ions) attached to it.

**Maltose Binding Protein (MBP)** – The *E. coli* MBP is a stable, soluble protein that binds maltose and amylose with high affinity. Expression of a peptide that contains both the MBP and the protein of interest results in a fusion protein that will probably bind to a column made from amylose cross-linked to agarose. This vastly simplifies purification of the heterologous protein. Of course, you then need to cleave the polypeptide to release your protein; this is usually (although not always) possible.

### **Problems with affinity chromatography**

Not all proteins/ligands can easily be linked to resins. The ligand must have some type of reactive group available.

Linking the ligand/protein may orient it in a position not favorable to interaction with your protein.

You must keep in mind that the column may act as an ion-exchanger or hydrophobic interaction column if your ligand/protein is charged or hydrophobic.

If your protein binds the column, you must also be able to elute from column (in other words, your protein may bind with such high affinity that you need to denature it to elute it from the column.

Affinity chromatography resins are expensive, and usually have somewhat limited capacity and very limited useful life-span. (Bacteria find most affinity columns to be admirable growth media, with consequent deleterious effects on the chromatographic usefulness of the columns.)

### **Differential Centrifugation**

Cells contain components of varying density. If the subcellular localization of a protein is known, it may be possible to partially purify the protein by removing subcellular fractions that do not contain the protein.

500 x g = whole cells, nuclei

10,000 x g mitochondria

100,000 x g microsomes (ER + Golgi)

Supernatant of 100,000 x g = cytosol

## **Ammonium Sulfate Precipitation**

In many cases, cell lysates can be loaded directly onto chromatography columns. However, in some cases other molecules present in the lysate interfere with binding of the protein to the resin. In addition, some resins (especially affinity resins and sepharose-based resins) are fairly expensive; loading crude cell lysates on these columns may result in binding of cellular material (*e.g.* lipids and DNA) that are difficult to remove, and which may damage the column. As a result, purification methods often begin with one of several possible simple techniques that remove at least some of these unwanted materials prior to using an expensive column.

One of the most commonly used crude purification techniques involves the use of differential **solubility**. Proteins exhibit differential solubility under a number of conditions. Unfortunately, adding an organic solvent, or drastically changing the pH, while capable of separating proteins by solubility, also tends to disrupt the three-dimensional structure of the protein. However, the use of ammonium sulfate will frequently result in differential precipitation of proteins that maintain the native folded structures.

Proteins precipitate with increasing ammonium sulfate concentrations, with most proteins precipitating somewhere between 10% and 60% ammonium sulfate. (The percentages are relative to a saturated solution, which has a concentration of about 4 M; thus most proteins precipitate between 0.4 M and 2.4 M.) This can allow a simple, partial, purification of a protein; if the protein of interest precipitates at 40% ammonium sulfate, many other proteins will remain in solution, as will many other non-protein molecules.

The mechanism of salting out is poorly understood. Increasing amounts of ammonium sulfate alter the structure of the solvent, and this, possibly in combination with other effects, induces aggregation of protein molecules.

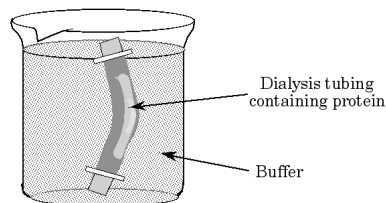
Most proteins are not damaged by ammonium sulfate precipitation, and can be resuspended in a small volume of buffer. Ammonium sulfate precipitation therefore has two roles: it is used to remove some types of impurities, and it is used to increase the concentration of the protein in solution. Ammonium sulfate precipitation results in a high salt concentration in the protein solution; this may be advantageous (if the next step is hydrophobic interaction chromatography), or deleterious (if the next step is ion exchange chromatography).

When necessary, two methods are frequently used to remove the salt. One method is gel filtration (salt is a small molecule, and separates easily from the protein on gel filtration columns). Other frequently used methods include dialysis and ultrafiltration.

## **Dialysis**

Dialysis involves placing the protein solution in a semi-permeable membrane (usually cellophane = cellulose acetate), and placing the membrane in a large container of buffer. Small molecules (such as salt ions) pass through the dialysis

membrane (moving from high concentration to low concentration), while large molecules are unable to cross the membrane. Dialysis membranes come in a variety of pore sizes, and are therefore useful for removing a variety of different sized solutes. In principle, dialysis could allow separation of large proteins from small ones; in practice, however, the pores in the tubing are insufficiently uniform to allow this technique to be used effectively.



The cellulose acetate tubing is usually shipped with various contaminants in the tubing. Preparation of the tubing for use with proteins involves placing the tubing in a buffer containing EDTA to chelate the metal ions. Boiling the tubing in the buffer accelerates the metal extraction and softens the tubing somewhat, which helps to prevent damage to the tubing during manipulation.

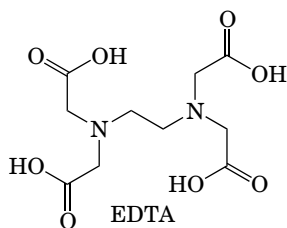
### Ultrafiltration

Like dialysis, ultrafiltration uses a semi-permeable membrane with pore sizes large enough to permit movement of small molecules and small enough to prevent the passage of proteins. However, unlike dialysis, where the small molecules move through the membrane by diffusion driven by their concentration gradient, and water tends to move inward driven by osmotic pressure, in ultrafiltration the small molecules move in aqueous solution because of an external driving force. Some ultrafiltration devices use elevated pressures (~500 kPa), usually applied by compressed nitrogen gas. Others use centrifugation where the centrifugal force induces the movement of water and small solutes through the membrane. Although ultrafiltration is somewhat less efficient at removing solutes, it is typically faster, and results in concentration of the protein sample.

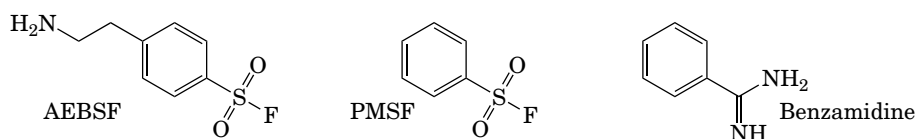
### Proteolysis and protease inhibition

**Buffer** – maintains pH and reduces activity of lysosomal proteases

**EDTA** – Ethylenediamine tetraacetic acid (EDTA) is a chelating agent; it is used to remove metal ions from solution. Some proteases are dependent on metal ions (especially calcium ions), so EDTA acts as an inhibitor of some proteases.



**PMSF** (phenylmethylsulfonyl fluoride), **AEBSF** 4-(2-aminoethylbenzenesulfonyl fluoride), and benzamidine inhibit serine proteases





**Pepstatin** – inhibits aspartyl proteases

**Leupeptin** – inhibits serine proteases and cysteine proteases

**Antipain** – inhibits papain, cysteine proteases, and serine proteases

### Stabilization

**Buffers** act to maintain pH. Some buffers interact chemically with proteins and may not be good choices for protein purification.

**NaCl** and **KCl** are usually used to maintain ionic strength.

**Sucrose** and **glucose** are used to stabilize membranes (especially those of subcellular organelles).

**Detergents** are used to extract and stabilize membrane proteins. In most cases, the concentration of detergent should be close to the CMC, so that the detergent molecules will self-associate. While non-ionic detergents are less likely to denature proteins than ionic detergents, the stability of a protein in a given detergent generally must be chosen empirically.

Glycerol is used to stabilize proteins, especially during freezing.

Dithiothreitol (DTT), Dithioerythritol (DTE), and mercaptoethanol are used to maintain cysteine side-chains in the reduced state.

### Electrophoretic techniques

Charged molecules exposed to an electric field are subjected to a force  $F = q\mathbf{E}$ , where  $q$  is the charge on the molecule, and  $\mathbf{E}$  is the electric field strength. The movement of the molecule is opposed by a frictional force. The steady-state velocity will occur when the magnitudes of the two forces are equal:

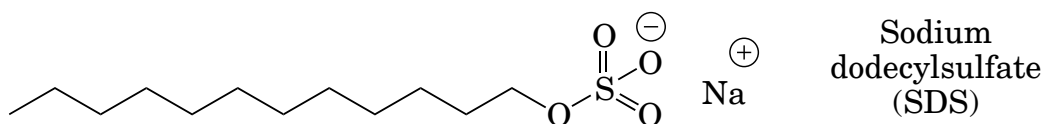
$$\mathbf{v} = q\mathbf{E}/f$$

This simple derivation ignores the complexity of an actual solution, where the presence of counterions alters the strength of the applied force in ways that have proven resistant to analysis. Thus, although the friction  $f$  is related to aspects of the molecular structure, little useful structural information can be from the movement of molecules in an electric field. In addition, although molecules move through solutions in response to electric fields, the separation achieved tends to be limited due to convection and diffusion phenomena.

Performing the electrophoresis in some type of buffer-permeated solid support greatly reduces the effects of convection and diffusion, and nearly all electrophoresis uses some form of solid support. Solid supports include paper and cellulose acetate, which generally do not interact with the solute, and gel matrices formed from polyacrylamide or agarose, which retard solute migration in a molecular size-dependent manner.

## SDS-PAGE

**SDS** (sodium dodecylsulfate, a compound also known as sodium lauryl sulfate) is a widely used detergent (for example, check the list of ingredients in your shampoo).

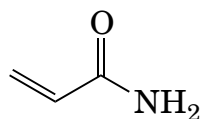


The hydrophobic tail of the SDS molecule interacts strongly with polypeptide chains. Each SDS molecule interacting with the polypeptide chain contributes a negative charge. Because SDS binds in large amounts (about 1.4 g of SDS per g of amino acid, corresponding to about one SDS for every 1.7 amino acid residues), the SDS charges overwhelm the intrinsic protein charge.

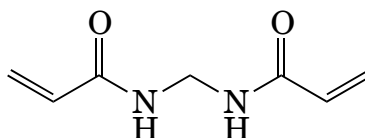
SDS also disrupts the hydrophobic core of the protein, and tends to cause proteins to unfold. The result of heating a protein in the presence of SDS is a rod-shaped detergent-coated molecule about 18 Å in diameter exhibiting a length proportional to the number of residues in the protein.

In most cases, the protein sample is exposed to both SDS and a reducing agent. The reducing agent cleaves disulfide bonds, and as a result assists in disruption of both tertiary and quaternary structure.

**PAGE** (polyacrylamide gel electrophoresis) is a technique for separating proteins based on charge and size. The gel limits the speed at which molecules can move, and prevents diffusion and convection from affecting the separation. Polymerization of acrylamide forms long polymers. The presence of *N,N'*-methylene bis-acrylamide results in the formation of cross-links between the polyacrylamide chains to create a matrix.

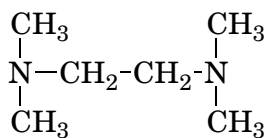


Acrylamide

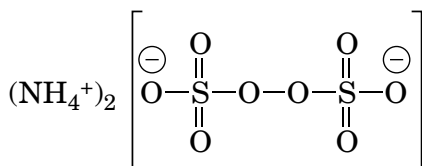


*N,N'*-Methylenebis(acrylamide)

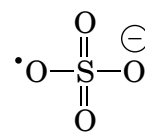
The polymerization reaction is a serial reaction using a free radical mechanism. The formation of the free radicals is initiated by the unstable compound ammonium persulfate. The sulfate radicals formed then react with *N,N,N',N'*-tetramethylethylenediamine (TEMED), forming TEMED radicals, which then react with acrylamide molecules to begin the actual polymerization reaction.



TEMED



Ammonium persulfate



Sulfate radical

Varying the amount of acrylamide monomer and bis-acrylamide cross-linker present controls the formation of the matrix. The use of larger amounts of these components results in a denser matrix. Denser matrices are used for separating smaller proteins; larger proteins may find the pores in a dense matrix too small to enter, and may therefore not enter the gel at all. The table below lists approximate useful ranges for different gel densities.

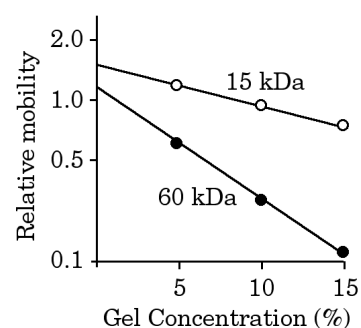
Percent acrylamide	Useful Molecular weight range
7	30,000 to 200,000
10	20,000 to 150,000
12	10,000 to 100,000
15	5,000 to 70,000

In SDS-PAGE, the protein molecules have approximately the same charge/mass ratio. Thus, charge is a linear function of size. In principle, all molecules should move at the same rate. In practice, the molecular migration is limited by the partition coefficient  $\sigma$ , because small molecules have a larger number of pores through the matrix they can use to migrate along the gel. Relative migration  $u$  is therefore related to both  $\sigma$  and molecular weight:

$$u = b - a \log M$$

As with gel filtration, a plot of  $\log(\text{molecular weight})$  versus migration distance for proteins of known size can be used as a standard curve to determine the molecular weight for other proteins.

Relative migration decreases as a function of gel concentration and as a function of size, with larger molecules moving much more slowly at higher gel concentrations. In addition to the percentage of acrylamide used, higher amounts of cross-linking will also make movement of large molecules difficult. (In most cases, the ratio of acrylamide to bisacrylamide is fixed, and only the acrylamide/bisacrylamide solution concentration is varied.)



For very large proteins (greater than 200 kDa) it is usually necessary to use agarose rather than polyacrylamide as the gel matrix, because agarose forms a lower density matrix.

### Isoelectric Focusing

IEF is another electrophoretic technique in which proteins separate according to their isoelectric points. In IEF, the gel contains ampholytes with differing pI values. (polymers containing both acidic and basic functional groups). The ampholytes migrate to their pI and act as buffers to maintain local pH. A series of appropriately

chosen ampholytes thus set up a gradient of pH values. Each protein then moves through the gel until the  $\text{pH} = \text{pI}$  for that protein, at which point it no longer has a net charge, and stops moving. Unlike SDS PAGE, IEF is run until the molecules reach equilibrium positions.

IEF can be combined with SDS PAGE. In IEF, proteins are separated based on their pI values. In SDS PAGE, proteins are separated because of differences in molecular weight. These properties are independent; and therefore running an IEF gel followed by SDS PAGE results in a two-dimensional separation.

### **Native gel electrophoresis**

In SDS-PAGE, the protein is denatured. Proteins are, however, charged molecules; it is possible to separate proteins on the bases of the intrinsic charge of the proteins. This method does **not** allow the measurement of the size of the protein; all proteins do not have identical charge/mass ratios. In addition, depending on the pH, some proteins will have positive charges, while others will have negative charges; this means that the proteins will migrate in opposite directions when subjected to the same electric field.

### **Ultracentrifugation**

Centrifugation methods separate molecules on the basis of their densities. The rate at which a molecule moves through the medium will be proportional to the size and shape of the molecule. Analytical ultracentrifugation requires special centrifuges and special rotors. It is time consuming (centrifuge runs often take days), and the technique rarely used.

### **Protein purification strategies**

Developing a scheme for purifying a protein remains an empirical process. However, in purifying a new protein, it is sometimes possible to adapt methods used for purifying similar proteins. In addition, planning the procedure before simply trying different methods can be extremely useful. Examples of this include using an ammonium sulfate precipitation step before a hydrophobic interaction chromatography step, because the high concentration of ammonium sulfate that results from the precipitation will allow the precipitated protein (or the non-precipitated protein remaining in solution) to be loaded directly onto the column. In contrast, an ammonium sulfate precipitated protein must be dialyzed (or otherwise desalted) prior to loading on an ion exchange column. Another frequently used scheme involves an inexpensive technique such as ammonium sulfate precipitation prior to remove bulk contaminants prior to running a higher resolution but more expensive technique such as affinity chromatography.

As with most scientific procedures, the more you know about the protein, and the more you know about protein purification, the more likely it is that you will be able to design a successful purification procedure.

Another important point to consider is that each step in a protein purification procedure is associated with costs. Three types of costs are important. One is that some reagents are expensive; incautious use of expensive reagents will waste them

(for example loading a crude homogenate directly on an expensive column will usually result in accumulation of large amounts of contaminants on the column that may prove difficult to remove). A second cost is time; more steps tend to take longer. A final cost is protein; each purification step involves some loss of the protein of interest. A purification scheme with too many steps will tend to have a low yield, and may result in purification of so little protein that the scheme must be repeated.

Obviously, if the protein of interest cannot be detected by some method, it will be impossible to determine whether a particular step resulted in purification of the protein. Purifying a protein is thus impossible if the protein cannot be detected, and therefore an assay of some kind is necessary.

## Summary

### **Proteins can be purified.**

Although all proteins are comprised of amino acids, the behavior of any given protein is entirely dependent on the precise composition and the precise order of the amino acids. By taking advantage of relatively minor differences in properties, it is quite possible to separate one protein from a mixture of proteins.

### **Most protein purification involves chromatography.**

Commonly used chromatographic methods include: **ion exchange** (which allows interaction with charged groups on the protein surface), **hydrophobic interaction** (which interacts with hydrophobic groups on the protein), **affinity** (which takes advantage of the ability of some proteins to bind other molecules with high affinity and specificity), and **gel filtration** (which allows the separation of molecules based on size).

### **A wide variety of methods are used for analyzing proteins.**

Some analytical methods in common use are gel filtration chromatography and SDS-PAGE, which both allow the measurement of protein size. Other electrophoretic methods allow the separation of proteins based on the intrinsic charge of the protein.